

# The Senso Question Answering System at QA@CLEF 2008

José Saias and Paulo Quaresma  
Departamento de Informática  
Universidade de Évora, Portugal  
{jsaias,pq}@di.uevora.pt

## Abstract

The University of Évora participation in QA@CLEF2008 was focused on the Portuguese monolingual task and was based on the updated SENSO Question Answering System.

This system uses a local knowledge base, providing semantic information for text search terms expansion. The solver module uses two components to collect plausible answers: the logic and the ad-hoc solvers. The logic solver starts by producing a First-Order Logic expression representing the question and a logic facts list representing the texts information and then it looks for answers within the facts list that unify and validate the question logic form. The ad-hoc solver is designed for cases where the answer can be directly detected in the text. Then all the results are merged for answer list validation, to filter and adjust answers weight.

The submitted run had only single answers (the system best answer). The overall accuracy was 46.5% and the overall Confidence Weighted Score was 0.23979. This paper has an overview of the system and its approach to QA@CLEF.

## Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.7 Digital Libraries; I.2 [Artificial Intelligence]: I.2.7 Natural Language Processing

## General Terms

Experimentation

## Keywords

Natural Language Processing, Question answering, Questions beyond factoids

## 1 Introduction

The Informatics Department of the University of Évora participated again<sup>1</sup> in the Question Answering (QA) task of the 2008's edition of Cross Language Evaluation Forum (CLEF)<sup>2</sup>. We registered for the Portuguese monolingual QA main task, testing our revised version of the SENSO QA System.

---

<sup>1</sup>The University of Évora has previous QA@CLEF participations in 2004 [1], 2005 [2] and last year [3]

<sup>2</sup>CLEF detailed information can be found at <http://www.clef-campaign.org/>

The system maintains the usage of a common sense knowledge base to assist on sentence analysis and text retrieval. The participation on QA@CLEF-2007 [3] gave some clues about improvements that could be done to reduce the errors. The text retrieval query generation process was updated to improve the candidate document selection. The solver module keeps the two components (logic and ad-hoc) that search in parallel for candidate answers, however there is a new component for answer validation. The question type is taken into account and a Web search might be performed to evaluate each answer and adjust its weight to a more reliable value.

The next section explains the system architecture. The followed methodology is described with examples in section 3. The evaluation of the obtained results is presented in section 4. Finally, some conclusions and future work are pointed out in section 5.

## 2 System Architecture

The main components are the same that were present in 2007. The updates occurred only inside some modules. Figure 1 illustrates the system major modules: *Libs*, *Query*, *Solver*, *Local KB* and *Web Interface*.

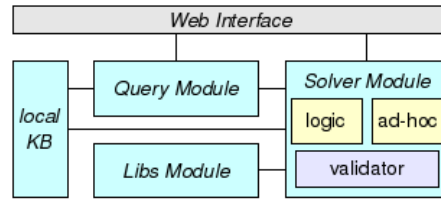


Figure 1: QA System Architecture

The *Query Module* performs the question analysis and selects a set of relevant documents for each question, as explained in section 3. The query group identifier determines if a query will be associated with the first from that group, to work with the same topic.

The *Libs Module* manages the text collections: news documents from Público and Folha de São Paulo from years 1994 and 1995, plus the Wikipedia documents. These collections are seen as libraries that contain the information needed for question answering. All texts are indexed with Lucene<sup>3</sup>, a full-featured text search engine, with a SENSO personalized stemmer for Portuguese.

The *Local KB* module has a starting knowledge base, containing common sense facts about places, entities and events. The information is available to a logic tool with some inference capabilities, that helps the automatic capture of sentence meaning.

Most of the changes happened in the *Solver Module*. It performs a search for plausible answers using two parallel approaches:

- *logic solver*: a logic-programming based tool that looks for answers to a question, being aware of the semantic expressed in *Local KB*
- *ad-hoc solver*: case-based answer detection for questions where the answer can be directly detected in the text

The results are merged, forming a global and weight sorted list of candidate answers. This list is then processed by the new component: the *answer validator*. Some answer values are refused if they are not in accordance with the question type. Besides filtering, each answer weight may suffer an adjustment to a more reliable value. The Web redundancy can be exploited as a method for answer validation in QA [6] [7]. The idea is to measure a statement popularity or acceptance with a Web search and take that into account for an answer accuracy validation.

<sup>3</sup>Apache Lucene is an open source project. <http://lucene.apache.org/>

The *Web Interface* layer allows an easier usage of the system. It is possible to consult text documents or check each intermediate step in the question analysis process and answer search. Next section explains the QA process.

### 3 Methodology

The SENSO approach for QA was firstly inspired on the authors previous work [4] and [5]. Appart from the improved operations on text retrieval query formulation and web answer validation, the process is quite similar to last year.

#### 3.1 Import the Text Collections

The XML collection files were processed and split into single texts, along with their metadata. The *Libs Module* keeps all these individual documents, being aware of their collection temporal context. Because we needed to perform some text search operations, the collections were indexed.

Each text was processed with PALAVRAS[8], a syntactical parser<sup>4</sup>. This tool is based on the Constraint Grammars formalism and gives a detailed text morpho-syntactical representation for later usage.

#### 3.2 Question Analysis

Each question is also processed with the syntactical parser PALAVRAS and a semantic analyzer able to obtain a partial semantic representation. The technique used for this process is based on Discourse Representation Structures (DRS) [9]. We are only dealing with a restricted semantic analysis and we are not able to handle every aspect of the semantics. The DRS is a First-Order Logic expression which the logic resolution tool will try to understand. The partial semantic representation of a sentence is a DRS built with two lists, one with the rewritten sentence and the other with the sentence discourse referents.

Let us consider the following QA@CLEF question:

*Que instrumento tocava Ringo Starr ? (What instrument did Ringo Starr play ?)*

The morpho-syntactical representation for the question, in figure 2, shows the parser tags identifying the subject, the predicate and the interrogative form *que* (*What*). Figure 3 has the DRS for the same question, with the semantic representation used by the system for later logic inference process. The system will search for something that *is an* instrument. At this point,

QUE:fc1	
=SUBJ:np	
==>N:pron-det('que' <interr> M/F S)	Que
==H:n('instrumento' <tool> <tool-mus> M S)	instrumento
=P:v-fin('tocar' IMPF 1/3S IND)	tocava
=ACC:prop('Ringo_Starr' <hum> M S)	Ringo_Starr
=	?

Figure 2: Syntactical Parser: sample output

a preliminary document retrieval task is done with Lucene, in order to elect a set of potentially relevant documents for each question. If no candidate documents are found the system cannot find an answer and the result is *NIL*.

The *Query Module* creates the Lucene search query. This is done with the question text terms and, for some, with their related terms as expressed in *Local KB*. So, when we are looking for *instrument* it is important to get documents that include synonyms or specialization terms (as

<sup>4</sup>Tool developed by Eckhard Bick. VISL Project: <http://visl.hum.sdu.dk/visl>

```

query(clef08qa0045,
  [ q(_220, 'que' , ['M/F', 'S', 'que' ],
    [ modif(nd,'instrumento', ['M','S','instrumento'] ) ] ),
    name(_221, 'Ringo_Starr' , ['M', 'S', 'Ringo_Starr' ],
      [ ] ),
    'tocar'(_220,_221,
      [ modif(verb,'tocar', ['IMPF','1/3S','IND'] ) ] ),
      [ ] ],
    [ ref(_220), ref(_221) ] ).

```

Figure 3: System’s logic representation for ‘What instrument did Ringo Starr play ?’

*piano, drums* and other), because those documents might be relevant as a possible answer source. When the question belongs to a cluster and it is not the first from that group, the query is fed with more terms in order to include the implicit topic. The system goes back to that cluster’s first question and gets their search terms and answer into the Lucene query.

### 3.3 Answer Engine

The *Solver Module* is responsible for finding a list of answers to a question, appreciate them and elect one as the best system answer. The list of candidate documents is processed in parallel by two different tools: the *logic solver* and the *ad-hoc solver*.

The semantic analyzer used before for the query will now create a DRS list for the selected texts. This list is a question dedicated Knowledge Base: the facts list. The *logic solver* is a logic-programming based module that performs a pragmatic interpretation of the query DRS over the full system knowledge base (the *Local KB* and the facts list). It tries to find the best explanations for the question logic form to be true. The inference process is done with the Prolog resolution algorithm.

The *ad-hoc solver* is a case-based answer selector for questions where the answer can be directly detected in the text. It uses a text pattern and structure based approach, explored before in [10]. The algorithm has several rules about the sentence structure for question and text, for certain question types. If the question is ‘What is X?’ and a text contains ‘X is a DEFINITION’, then *DEFINITION* is a possible answer.

A concrete example for this situation is the question:

*O que é uma cítara ? (What is a cítara ?)*

There was a Wikipedia text stating:

*A ‘cítara’ é um instrumento musical de várias cordas presas sobre um arco de madeira, com ou sem caixa de ressonância, que se tocavam com ambas as mãos.*

This direct match gave one candidate answer, defining a cítara as a musical instrument having strings attached to a wood arc.

Each candidate answer has a weight and a snippet: sentence or expression justifying the answer and its document identifier. The results are merged, forming a global and weight sorted list that is processed by the *answer validator*. This new component does an appreciation of each answer value. Some might be refused if they are not in accordance with the question type.

Based on the principle that the Web redundancy can be exploited as a method for answer validation in question answering [6] [7], the system does a quick Web search to measure the answer value popularity with respect to the question. As a consequence, each answer weight may

suffer an adjustment to a more reliable value.

If there is no candidate answer the system returns *NIL* as result. When the system finds more than one answer for a question, then the QA@CLEF answer is the one with the highest weight.

## 4 Results

As before, the University of Évora’s group QA@CLEF participation was focused on the monolingual Portuguese task. The set of 200 questions was processed by the SENSO QA system and one run output was sent for evaluation. The system answers were classified as *Right* for 93 questions, which corresponds to an overall accuracy score of 46.50% (4.5% more than obtained last year).

The system returned only 21 *NIL* answers, significantly less when comparing to 2007. Despite this appears to be a better value, the accuracy for the *NIL* question type went down from 10.81% to 9.52%. In the *Factoids* category the system had an accuracy of 40.74%, quite similar to last year. The best relative accuracy result was again on the *Definition* question type with 85.71%, which represents an improvement since 2007.

Table 1 shows the system detailed results, specifying the accuracy values per question type.

Question Type	#	Right	Wrong	Unsupported	Inexact	Accuracy
Nil	21 returned	2	19	0	0	9.52%
Temporally Restricted	16	4	12	0	0	25.00%
Definition	28	24	2	0	2	85.71%
Lists	10	3	5	0	2	30.00%
Factoids	162	66	87	2	7	40.74%
<b>All Questions</b>	<b>200</b>	<b>93</b>	<b>94</b>	<b>2</b>	<b>11</b>	<b>46.50%</b>

Table 1: University of Évora’s results in QA@CLEF2008

The overall Confidence Weighted Score over all assessed questions is  $47.959/200 = 0.23979$ . The overall accuracy is slightly better than the obtained last year. The next section has some conclusions about this participation.

## 5 Discussion

The University of Évora’s participation on QA@CLEF 2008 was based on their SENSO Question Answering System. Being the second time this QA system is used, the results are in line with the expected: close to or a little better than last year.

The document retrieval process was updated. One of the changes was in the Portuguese stemmer used in the Lucene indexing and search operations. This and the text search query generation process led to the identification of more candidate documents, decreasing the number of *NIL* answers.

The *logic solver* has a few problems with DRS generation, specially when analyzing morpho-syntactical representation of non-trivial sentences. Other problems happen in the pragmatic interpretation of the DRS. Most of the answers found by the system are from the *ad-hoc solver*.

The answer validation process gave a relevant contribution to accuracy. The redundancy of Web information allows the usage of a search result as a clue for the validity of an answer.

The *ad-hoc solver* is a rule based answer generator. Some of those rules need an adjustment. Before applying our system to other source languages, there is some work to do in order to make the system components independent from the language. The rules in *ad-hoc solver* are language dependent. One possibility for a future participation is the submission of multiple answers per question. That can be accomplished with this system by selecting the *N* most weighted answers.

## References

- [1] Paulo Quaresma, Luis Quintano, Irene Rodrigues, José Saias and Pedro Salgueiro. *The University of Évora approach to QA@CLEF-2004*. CLEF 2004 Working Notes.
- [2] Paulo Quaresma and Irene Rodrigues. *A Logic Programming Based Approach To QA@CLEF05 Track*. CLEF 2005 Working Notes.
- [3] José Saias and Paulo Quaresma. The senso question answering approach to portuguese qa@clef-2007. Technical report, CLEF 2007 Working Notes, Cross-Language Evaluation Forum Workshop, Budapest, Hungary, 2007. ISSN: 1818-8044, ISBN: 2-912335-32-9.
- [4] José Saias and Paulo Quaresma. *A methodology to create ontology-based information retrieval systems*. Fernando Moura Pires and Salvador Abreu (Eds): Progress in Artificial Intelligence - Proceedings of the 11th Portuguese Conference on Artificial Intelligence, EPIA'03, Beja, Portugal, 2003. Springer-Verlag, ISBN: 3-540-20589-6
- [5] José Saias and Paulo Quaresma. *A proposal for an ontology supported news reader and question-answer system*. Solange Oliveira Rezende et al. (Eds): 2nd Workshop on Ontologies and their Applications (WONTO'06) in the Proceedings of International Joint Conference, 10th IBERAMIA, ICMC-USP, Ribeirão Preto, Brazil, 2006. ISBN: 85-87837-11-7.
- [6] Charles Clarke, Gordon Cormack and Thomas Lynam *Exploiting redundancy in question answering* In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2001), 2001. ISBN 1-58113-331-6
- [7] Bernardo Magnini, Matteo Negri, Roberto Prevete and Hristo Tanev *Is It the Right Answer? Exploiting Web Redundancy for Answer Validation* In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, 2002 DOI: 10.3115/1073083.1073154
- [8] Eckhard Bick. *The Parsing System "Palavras". Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press, 2000.
- [9] Kamp, H. and Reyle, U. *From Discourse to Logic*. Kluwer: Dordrecht. 1993
- [10] Hristo Tanev. *Extraction of Definitions for Bulgarian*. CLEF 2006 Working Notes.